Reprinted from AGRONOMY JOURNAL
Vol. 66, March-April 1974, p. 259-262

Smoothing Data with Fourier Transformations[1]

B. A. Kimball[2]

ABSTRACT

The natural variability of agronomic data often masks underlying true curves. A relatively new Fourier transform data smoothing technique, when tested on several sets of noisy agronomic data, gave very satisfactory reductions of this variability. The data were smoothed by computing the Fourier transform, setting high frequency noise components of the resulting variance spectrum to zero, and then computing the inverse Fourier transform. The key to the technique's success lies in the fact that the data were dominated by relatively low frequencies of variation which could be separated from the higher frequency noise. This simple, lucid interpretation of the filtering action makes Fourier transform smoothing even more attractive than other methods whose filtering action is not so intuitively obvious. The technique was superior to two other methods in extracting a true curve from noisy artificial data.

Additional index words: **Curve fitting, Digital filtering, Digital smoothing, Spectral analysis, Power spectrum, Variance spectrum.**

THE environmental factors or parameters that affect plant growth possess much natural variability, and the evaluation of such parameters involves significant sampling and/or measurement errors. For many of these parameters, agronomists are able to utilize standard statistical methods of data analysis to deduce reliable estimates of the true parameter values. However, some of the parameters are sampled as time series, a class of data whose analysis has received relatively less attention in agronomic literature.

A time series is a set of observations of some parameter, usually taken at equal intervals of time. The usual method for obtaining estimates of the true values of the parameters has been to smooth the data by computing weighted running averages of the original observations. Recently Hayes et al. (1973) published a smoothing technique based on the Fourier transform which they found would smooth electroanalytical data well. The purpose of this paper is to illustrate the use of this technique for smoothing several agronomic variables and to suggest that it is an attractive alternative to running averages and other methods.

THEORY OF FOURIER TRANSFORMS AND SPECTRAL ANALYSIS

The Fourier transform of a set of data points, $x(t_k)$, is defined by the equation

$$R(f_n) + i\,I(f_n) = \Delta t \sum_{k=0}^{N-1} x(t_k)\left[\cos(2\pi f_n t_k) - i\,\sin(2\pi f_n t_k)\right] \qquad [1]$$

$$n = 0, \pm 1, \ldots \pm N/2$$

where:

x = variable under study
t = time
t_k = kth time = $k\Delta t$ (the first time is defined as t_0)
$x(t_k)$ = value of x at the kth time
N = total number of observation points
Δt = time interval between data points
f_n = nth frequency = n/T
T = total time period = $N\Delta t$
$R(f_n)$ = real part of Fourier coefficient for nth frequency
$I(f_n)$ = imaginary part of Fourier coefficient for nth frequency
i = $(-1)^{1/2}$

This transform has often been used in the statistical technique known as spectral analysis because

$$s(f_n) = 2\,\frac{\left\{[R(f_n)]^2 + [I(f_n)]^2\right\}}{N\Delta t} \qquad [2]$$

where $s(f_n)$ is the value of the spectral density function corresponding to the nth frequency. Spectral analysis is used to evaluate the contributions of different frequencies of fluctuations to the total variance of an entity, such as air pressure (Kimball and Lemon, 1970), which changes with time. A graph of $s(f)$ plotted against frequency is known as a variance spectrum or power spectrum. The area under the curve equals the total variance, and the height of any portion of the curve illustrates the contribution of that corresponding frequency band to the variance. For example, Fig. 1A shows the diurnal variation of the gravimetric soil moisture content for the 0 to 5 mm layer of bare Avondale clay loam (formerly Adelanto loam) 7 days after irrigation (Jackson, 1973), and Fig. 1C is its variance

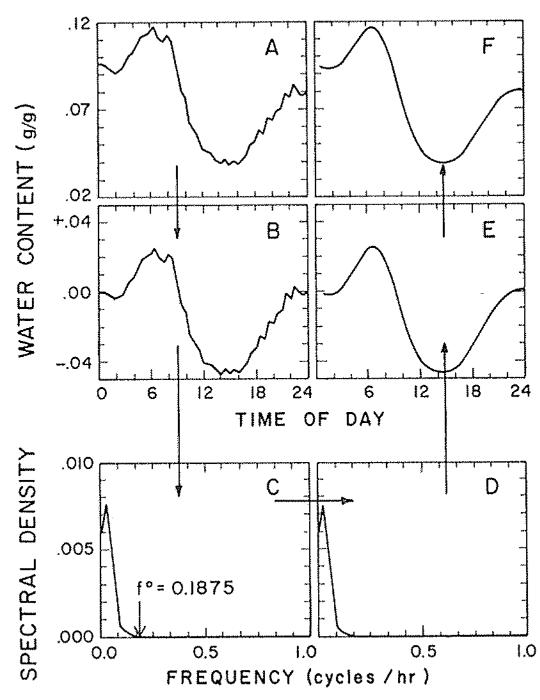

Fig. 1. Illustration of the steps involved with Fourier transform smoothing: A → B translation-rotation, B → C Fourier transform, C → D removal of noise portion of spectrum, D → E inverse Fourier transform, E → F inverse translation-rotation. [Water content data from Jackson (1973)].

[1] Contribution of the Agricultural Research Service, USDA. Received Aug. 10, 1973.
[2] Soil Scientist, U. S. Water Conservation Laboratory, Phoenix, AZ 85040.

spectrum. Since the spectral density in Fig. 1C becomes very small for frequencies greater than 0.1875 cycle/hour (=1 cycle/5.333 hour), most of the variance in moisture content was caused by frequencies lower than 0.1875. More information about spectral analysis can be found in Blackman and Tuckey (1958), Panofsky and Brier (1958), Lumley and Panofsky (1964), Jenkens and Watts (1968), and Kimball (1969, 1970).

The complement of Equation 1 is the inverse Fourier transform defined by

$$x(t_k) = \frac{1}{N\Delta t} \sum_{n=-N/2}^{n=+N/2} [R(f_n)+i\,I(f_n)][\cos(2\pi f_n t_k) + i\,\sin(2\pi f_n t_k)] \qquad [3]$$

$$k = 0, 1, 2 \ldots N-1$$

Using Equation [3], the original data are restored by an inverse Fourier transform of the real and imaginary Fourier coefficients calculated from Equation [1].

Fourier transforms can be computed using programs based directly on Equations [1] and [3]. However, this procedure is slow because there are so many complex multiplications and additions. Cooley and Tukey (1965) have presented a new algorithm for the fast computation of Fourier transforms. Their algorithm, known as the FFT (fast Fourier transform), has had wide application in communications engineering. More information may be obtained in the June 1967 issue of IEEE Transactions on Audio and Electroacoustics, Vol. 15, No. 2, which was devoted entirely to the FFT and how it relates to spectral analysis and other subjects. An easily understood description of the FFT is given by Brigham and Morrow (1967). Following their outline, the author has programmed the fast Fourier transform in BASIC programming language.

The fast Fourier transform algorithm owes its speed to a dramatic reduction in the required number of multiplications by factorization of a weighting matrix. The factorization is possible only when the number of data points is chosen to be a power of two, but this restriction is not serious because the user can simply add zeros at the end of his data set until the number of data points plus zeros equals a power of two. The N in Equation [1] becomes this total. The procedure is valid because the addition of zeros does not change the variance of the data set.

The key to the smoothing of data by the method of Fourier transformations involves manipulating the variance spectrum prior to computing an inverse transform. For many cases of agronomic data, the variance spectrum can be broken into two parts. In Fig. 1C, the portion to the left of the arrow represents the very important low frequency components. The portion to the right of the arrow has amplitudes so small they are not visible on the graph, yet this right portion of the graph represents the rather significant presence of high frequency noise in the original data. By deliberately setting the $R(f_n)$ and $I(f_n)$ equal to zero for all frequencies to the right of the arrow [a procedure first suggested by Morrison (1963)], and then calculating the inverse Fourier transform, the smooth data of Fig. 1F were obtained. The frequency designated by the arrow is the cutoff frequency, $f°$.

The straightforward application of the transform procedure outlined above will be unsatisfactory, however, for most agronomic data. As discussed by Hayes et al. (1973), the data must begin and end with zero or near zero values or spurious peaks will be introduced. To circumvent this problem, they suggest a rotation-translation as illustrated in Fig. 1A and 1B, and as defined by

$$x'(t_k) = x(t_k) - \left[x(t_0) + \frac{[x(t_{N-1}) - x(t_0)]k}{(N-1)} \right] \qquad [4]$$

$$k = 0, 1, 2 \ldots N-1$$

prior to computing the Fourier transform. After setting the high frequency noise portion of the spectrum equal to zero and computing the inverse transform, the smooth data can be restored by the inverse of Equation [4] (i.e., change the minus outside the brackets to a plus).

As discussed by Hayes et al., however, the translocation-rotation transformation can lead to erroneous results if the data are significantly inaccurate around the initiation and termination points. At our laboratory we have obtained satisfactory results by obtaining and including extra data before and after the time period of particular interest. Where this remedy is not possible, a weighted average of several initial and of several terminal points substituted for the initial and terminal points respectively probably would give satisfactory results.

The steps involved in Fourier transform smoothing can be summarized as follows:

1) Rotate-translate using Equation [4], Fig. 1A $\rightarrow$ 1B;
2) Add zeros if necessary to use the FFT algorithm;
3) Fourier transform using Equation [1], Fig. 1B $\rightarrow$ 1C;
4) Set $R(f_n)$ and $I(f_n) = 0$ for $f_n \geqslant f°$, Fig. 1C $\rightarrow$ 1D;
5) Inverse Fourier transform using Equation [3], Fig. 1D $\rightarrow$ 1E;
6) Inverse rotate-translate using the inverse of Equation [4], Fig. 1E $\rightarrow$ 1F.

APPLICATIONS

Some additional examples of Fourier transform smoothing are shown in Fig. 2 and 3. In Fig. 2 the output from the load cell of a lysimeter is shown plotted against time of day along with the curve obtained by the Fourier transform smoothing. Wind gusts and other factors caused the lysimeter to bounce somewhat and produce the irregularities in the output. Fourier transform smoothing of the data prior to taking the differences has proved to be a very satisfactory method for extracting evaporation rates. In Fig. 3, the chloride content of the upper 5 mm of bare Avondale clay loam is shown plotted against time of day for days 3 and 7 after irrigation [from Nakayama et al. (1973)]. Fourier transform smoothing produced the smooth curves from the noisy data.

The Fourier transform smoothing technique utilizes one arbitrary parameter, the cutoff frequency, which must be specified by the user. The particular value which is appropriate must be high enough so that none of the components characterizing the data are lost, but low enough so that as much noise as possible is removed. Therefore, the proper choice depends upon prior knowledge or intuition on the part of the experimenter about the maximum frequency with which it is physically possible for his system to vary. Although this Fourier transform

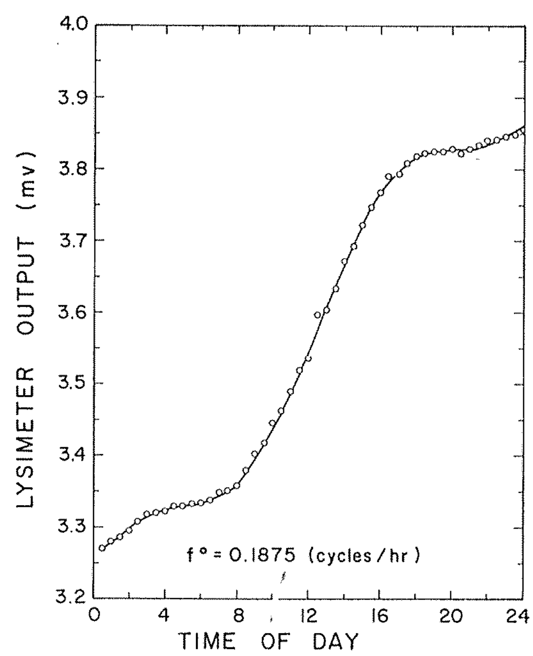

Fig. 2. Fourier transform smoothing of lysimeter data.

method of smoothing does contain this element of arbitrariness, it is present in no greater degree than with other methods. Indeed, the ability to specify a definite cutoff frequency with Fourier transform smoothing lends itself to a more lucid interpretation of the nature of the filtering action obtained by smoothing.

The effect of changing the cutoff frequency upon the smoothing of water content data is illustrated in Fig. 4. In Fig. 4A the jagged curves are the same water content data as in Fig. 1A replotted three times with the ordinate shifted 0.02 units between each replot. The arrows in Fig. 4B show the cutoff frequency used for each of three computer runs, and the smooth curves superimposed on Fig. 4A are the smoothed data computed using the respective cutoff frequencies. The upper two curves both follow the actual data very well and would probably give satisfactory results. The major difference occurs in the shape of the peak at 0600 hours, the curve for $f^\circ = 0.34375$ cycle/hour being flatter. The curve for $f^\circ = 0.09375$ does not follow the data adequately. This is not surprising because the arrow in Fig. 4B shows that significant components from the shoulder of the peak in the variance spectrum were cut off.

The ability of Fourier transform smoothing to extract a true curve from noisy data was assessed as follows. Data points were computed for $t = .5, 1, 1.5, 2 \ldots 24$ from the equation

$$W = 0.09 + (-0.02/24)t + 0.03 \sin(2\pi t/24) \quad [5]$$

which somewhat simulates the water content near the surface of bare Avondale clay loam several days after irrigation. The equation describes a sine wave of period 24 hours with an amplitude of 0.03 g/g superimposed on a linear drying trend of 0.02 g/g per 24 hours. Then five sets of normally distributed "noise" of standard deviation approximately 0.005 g/g was generated using a random number generator and a discretized normal distribution curve. The five sets of noise were added to the "true" data points and the resulting noisy curves were smoothed using the Fourier transform smoothing technique as well as others.

Figure 5 illustrates the results obtained from the first set of generated data for $f^\circ = 0.1875$. The dashed

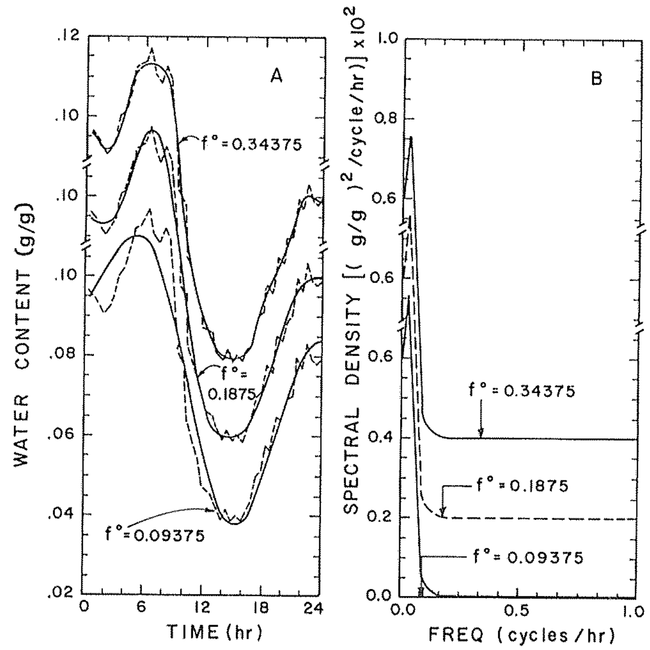

Fig. 4. Effect of three cutoff frequencies on the Fourier transform smoothing of soil water content data. [Data from Jackson (1973)].

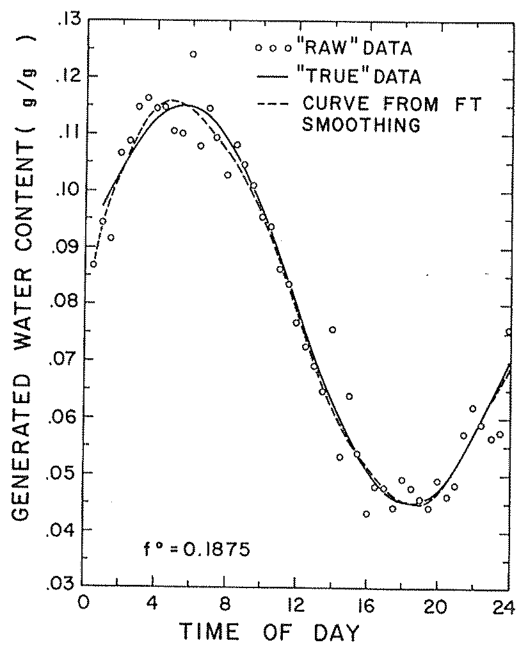

Fig. 5. Illustration of the ability of Fourier transform smoothing to extract approximately the true curve from noisy artificial data. The true curve is actually a simulation of surface soil moisture content with a sine wave superimposed on a linear drying trend. The raw artificial data were generated by adding normally distributed noise to the data points of the true curve.

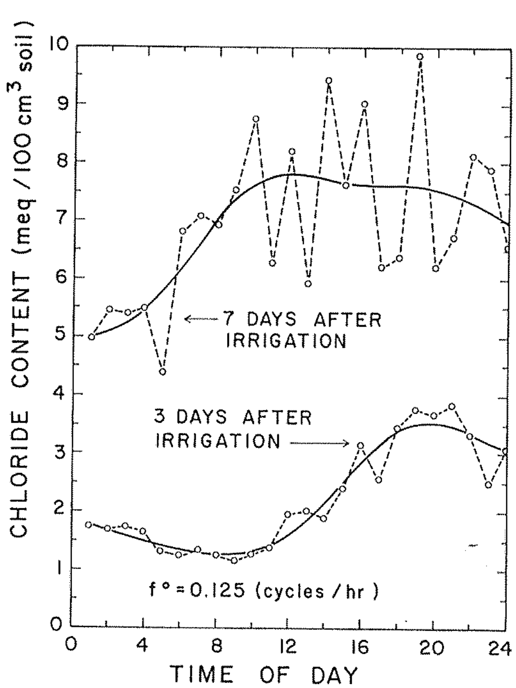

Fig. 3. Fourier transform smoothing of chloride salinity data. [Data from Nakayama et al. (1973)].

Table 1. Standard deviations of five sets of generated water contents and the reductions due to three methods of smoothing.

	Standard Deviation, g/g				
Data set	Raw	1, 2, 3, 2, 1 Weighted running average	Parabolic splines*	Fourier transform $f^0 = .1875$	$f^0 = .125$
1	.0046	.0019	.0013	.0011	.0012
2	.0064	.0024	.0017	.0017	.0016
3	.0058	.0024	.0017	.0017	.0015
4	.0049	.0022	.0025	.0020	.0019
5	.0044	.0021	.0020	.0016	.0014
Average reduction, %		57.8	64.2	68.8	70.4

* From DuChateau et al. (1972).

curve obtained by Fourier transform smoothing of the "raw" data points approximates the solid "true" curve, Equation [5], well. The ability of Fourier transform smoothing and other smoothing techniques to reduce the standard deviation of the five generated water content data sets is shown in Table 1. Here the standard deviation was computed as the square root of the sum of the differences between estimated value and the true values from Equation [5]. Smoothing by the Fourier transform method reduced the standard deviation by an average of about 70% from the original standard deviation of the raw data for both cutoff frequencies used. The more familiar 1, 2, 3, 2, 1 weighted running average technique [eg. as used by Jackson (1973)] only achieved a reduction of 58%. The results obtained by the parabolic splines technique presented by DuChateau et al. (1972) were intermediate with an average reduction of 64%.

Obviously for this artificial data, the Fourier transform was superior, but all three methods achieved large reductions in the amount of scatter and all would be satisfactory for many cases. The simplicity of the weighted running average makes it attractive to users with limited computing facilities. The parabolic splines technique (DuChateau et al., (1972) has the advantage that it can be used with unequal intervals, and also that interpolations of function values or slopes can be made from quadratic functions rather than connection lines between discrete points (although running quadratics could be fitted to the smoothed points of the other methods too, if desired).

The choice of method also depends upon what is known about the variable to be smoothed. The artificial data used here was generated by the use of a function which had a frequency of 1/24 hours and consequently, removing noise due to frequencies higher than 1/8 hours (= 0.125 cycle/hour) had to improve the situation. On the other hand, if some variable is known to change both rapidly and slowly, such as $X^{1/2}$ in the example used by DuChateau et al. (1972), then both low and high frequency components are needed to describe the function, and the highs should not be removed. Therefore, for such data the parabolic splines function can be expected to give more satisfactory results. However, it appears that many things in nature which command the attention of agronomists (three examples are presented here) are dominated by relatively slow changes, and frequent sampling coupled with Fourier transform smoothing will produce very satisfactory results. Moreover, the simple interpretation of the Fourier transform smoothing filter — that it is a simple separation of high from low frequencies of variation — makes Fourier transform smoothing an attractive alternative to the other methods whose filtering action is not so intuitively obvious.

ACKNOWLEDGMENT

I gratefully acknowledge the contribution of Dr. F. S. Nakayama who called the pertinent chemical literature to my attention.

LITERATURE CITED

Blackman, R. B., and J. W. Tukey. 1958. The measurement of power spectra. Dover Publications, New York. 190 p.

Brigham, E. O., and R. E. Morrow. 1967. The fast Fourier transform. IEEE Spectrum 4 (3):63-70.

Cooley, J. W., and J. W. Tukey. 1965. An algorithm for the machine calculation of complex Fourier series. Math. Comput. 19:297-301.

DuChateau, P. C., D. L. Nofziger, L. R. Ahuia, and D. Swartzendruber. 1972. Experimental curves and rates of change from piecewise parabolic fits. Agron. J. 64:538-542.

Hayes, J. W., D. E. Glover, D. E. Smith, and M. W. Overton. 1973. Some observations on digital smoothing of electroanalytical data based on the Fourier transformation. Anal. Chem. 45:277-284.

Jackson, R. D. 1973. Diurnal changes in soil water content during drying. p. 37-55. *In* R. R. Bruce, K. W. Flach, and H. M. Taylor (ed.) Field soil water regime. Spec. Publ. No. 5, Soil Sci. Soc. Amer., Madison, Wis.

Jenkins, G. M., and D. G. Watts. 1968. Spectral analysis and its applications. Holden-Day, San Francisco. 525 p.

Kimball, B. A. 1969. The energy budget at the earth's surface: Basic concepts of spectral analysis by digital means. Microclimate investigations interim report 69-2, USDA and Cornell University. 63 p.

———. 1970. Effects of air turbulence upon gas exchange from soil. Ph.D. Thesis. Cornell Univ. (No. 70-14, 391). Univ. Microfilms. Ann Arbor, Mich. (Diss. Abstr. 31b (3):1008b).

———, and E. R. Lemon. 1970. Spectra of air pressure fluctuations at the soil surface. J. Geophys. Res. 75:6771-6777.

Lumley, J. S., and H. A. Panofsky. 1964. The structure of atmospheric turbulence. John Wiley and Sons, Inc., New York. 239 p.

Morrison, J. D. 1963. On the optimum use of ionization-efficiency data. J. Chem. Phys. 390:200-207.

Nakayama, F. S., R. D. Jackson, B. A. Kimball, and R. J. Reginato. 1973. Diurnal soil-water evaporation: chloride movement and accumulation near the soil surface. Soil Sci. Soc. Amer. Proc. 37:509-513.

Panofsky, H. A., and G. W. Brier. 1958. Some applications of statistics to meteorology. The Pennsylvania State University Press, University Park, Pa. 223 p.